

# Bootstrapping the Kaplan-Meier Estimator on the Whole Line

Dennis Dobler\*

May 19, 2016

## Abstract

This article is concerned with proving the consistency of Efron's (1981) bootstrap for the Kaplan-Meier estimator on the whole support of a survival function. While other works address the asymptotic Gaussianity of the estimator itself without restricting time (e.g. Gill, 1983, and Ying, 1989), we enable the construction of bootstrap-based time-simultaneous confidence bands for the whole survival function. Other practical applications include bootstrap-based confidence bands for the mean residual life-time function or the Lorenz curve as well as confidence intervals for the Gini index.

**Keywords:** Counting process; Right-censoring; Resampling; Efron's bootstrap; Mean-residual lifetime; Lorenz curve; Gini index.

\* University of Ulm, Institute of Statistics, Germany

# 1 Introduction

This article reconsiders Efron's (1981) bootstrap of Kaplan-Meier estimators. It is well-known that drawing with replacement directly from the original observations consisting of (event time, censoring indicator) reproduces the correct covariance structure; see e.g. Akritas (1986), Lo and Singh (1986), Horvath and Yandell (1987) or van der Vaart and Wellner (1996) for an application in empirical processes. Let  $T : \Omega \rightarrow (0, \tau)$  be a continuously distributed random survival time with survival function given by  $S(t) = 1 - F(t) = P(T > t)$ . For conceptual convenience we mainly refer to  $T$  as a random *survival time*, although other interpretations are also reasonable; see the examples below. In the previously mentioned articles the typical assumption  $S(\tau) > 0$  is met for mathematical convenience in proving weak convergence of estimators for  $S$  on the Skorohod space  $D[0, \tau]$  and because most studies involve a rather strict censoring mechanism: after a pre-specified end of study time each individual without an observed event is considered as right-censored. Thus, it is often not possible to draw inference on functionals of the whole survival function.

Some functionals, however, indeed require the possibility to observe arbitrarily large survival times. For instance, consider the *mean residual life-time* function

$$t \mapsto g(t) = \mathbb{E}[T - t \mid T > t] = \frac{1}{S(t)} \int_t^\tau S(u) du; \quad (1.1)$$

see e.g. Meilijson (1972), Gill (1983), Remark 3.3, and Stute and Wang (1993). This function describes the expected remaining life-time given the survival until a point of time  $t > 0$ . Another, econometric example of a functional of whole survival curves is the *Lorenz curve*

$$p \mapsto L(p) = \mu^{-1} \int_0^p F^{-1}(t) dt = \mu^{-1} \int_0^{F^{-1}(p)} s dF(s), \quad (1.2)$$

where  $F^{-1}(t) = \inf\{u \geq 0 : F(u) \geq t\}$  is the left-continuous generalized inverse of  $F$  and  $\mu = \int_0^\infty t dF(t)$  is its mean. With the interpretation of  $T$  being the income of a random individual in a population, this function  $L$  obviously represents the total income of the lowest  $p$ th fraction of all incomes. A closely related quantity is the *Gini index*

$$G = \frac{\int_0^1 (u - L(u)) du}{\int_0^1 u du} \in [0, 1] \quad (1.3)$$

as a measure of uniformity of all incomes within a population; see e.g. Tse (2006). The value  $G = 0$  represents perfect equality of all incomes, whereas  $G = 1$  describes the other extreme: only one persons gains everything and the rest nothing.

All quantities (1.1), (1.2) and (1.3) are statistical functionals of the *whole* survival function  $S$ . First analyzing  $S$  only on a subset of its support results inevitably in an alternation of the above functionals in a second step. And this affects the interpretation of such quantities. In order to circumvent such problems, estimating the whole survival function is the obvious solution: Henceforth, denote by  $\tau = \inf\{t \geq 0 : S(t) = 0\} \in (0, \infty]$  the support's right end point. Wang (1987) and Stute and Wang (1993) showed the uniform consistency of the *Kaplan-Meier* or *product-limit estimator*  $(\hat{S}(t))_{t \in [0, \tau]}$  for  $(S(t))_{t \in [0, \tau]}$  and Gill (1983) and Ying (1989) proved its weak convergence on the Skorohod space  $D[0, \tau]$ . For robust statistical inference procedures concerning the above functionals of  $S$  it is thus necessary to extend well-known bootstrap results for the Kaplan-Meier estimator to the whole Skorohod space  $D[0, \tau]$ . After presenting this primary result we deduce inference procedures for the quantities (1.1) to (1.3).

This article is organized as follows. Section 2 introduces all required estimators, recapitulates previous weak convergence results on  $D[0, t]$ ,  $t < \tau$ , and provides handy results for checking all main assumptions. The main theorems on weak convergence of the bootstrap Kaplan-Meier estimator are presented in Section 3, including a consistency theorem for a bootstrap variance function estimator. Section 4 deduces inference procedures for (1.1) to (1.3) and the final Section 5 gives a discussion on future research possibilities. All proofs are given in the Appendix. Most of this article's results originate from the University of Ulm PhD thesis of Dennis Dobler; cf. Chapter 6 and 7 of Dobler (2016).

## 2 Preliminary Results

Let  $T_1, \dots, T_n : \Omega \rightarrow (0, \infty)$ ,  $n \in \mathbb{N}$ , be independent survival times with continuous survival functions  $S(t) = 1 - F(t) = P(T_1 > t)$  and cumulative hazard function  $A(t) = \int_0^t \alpha(u) du = -\int_0^t (dS)/S_- = -\log S(t)$ . Independent thereof, let  $C_1, \dots, C_n : \Omega \rightarrow (0, \infty)$  be i.i.d. (censoring) random variables with (possibly discontinuous) survival function  $G(t) = P(C_1 > t)$  such that the observable data consist of all  $1 \leq i \leq n$  pairs  $(X_i, \delta_i) := (T_i \wedge C_i, \mathbf{1}\{X_i = T_i\})$ . Here,  $\mathbf{1}\{\cdot\}$  is the indicator function. Thus, the survival function of  $X_1$  is  $H = S \cdot G$ . The Kaplan-Meier estimator is defined by  $\hat{S}_n(t) = \prod_{i: X_{i:n} \leq t} (1 - \frac{\delta_{[i:n]}}{n-i+1})$ , where  $(X_{1:n}, \dots, X_{n:n})$  is the order statistic of  $(X_1, \dots, X_n)$  and  $(\delta_{[1:n]}, \dots, \delta_{[n:n]})$  are their concomitant censoring indicators. Throughout, we assume that

$$-\int_0^\tau \frac{dS}{G_-} < \infty \quad (2.1)$$

which restricts the magnitude of censoring to a reasonable level. For instance Gill (1983), Ying (1989) and Akritas and Brunner (1997) require this condition for an analysis of the large sample properties of Kaplan-Meier estimators on the whole support  $[0, \tau]$ . Thereof, Gill (1983) requires Condition (2.1) for a vanishing upper bound in Lengart's inequality. Obviously, the above condition implies that  $[0, \tau]$  is contained in the support of  $G$ ; see also Allignol et al. (2014) for a similar condition in a non-Markov illness-death model, reduced to a competing risks problem.

Denote by  $\hat{T}_n := \max_{i \leq n} X_i$  the largest observed event or censoring time and let, for a function  $t \mapsto f(t)$ , the notation  $f^{\hat{T}_n}$  be its stopped version, i.e.,  $f^{\hat{T}_n}(t) = f(t \wedge \hat{T}_n)$ . The monotone function  $t \mapsto \sigma^2(t) = \int_0^t (dA)/H_-$  is the asymptotic variance function of the related Nelson-Aalen estimator for  $A$  and reappears in the asymptotic covariance function of  $\hat{S}_n$ . Throughout, all convergences (in distribution, probability, or almost surely) are understood to hold as  $n \rightarrow \infty$  and convergence in distribution and in probability are denoted by  $\xrightarrow{d}$  and  $\xrightarrow{p}$ , respectively. The present theory relies on the following weak convergence results for the Kaplan-Meier process  $\hat{S}_n$  of  $S$ .

LEMMA 2.1. *Let  $B$  denote a Brownian motion on  $[0, \tau]$  and suppose (2.1) holds.*

- (a) Theorem 1.2(i) of (Gill, 1983): *On  $D[0, \tau]$  we have  $\sqrt{n}(\hat{S}_n - S)^{\hat{T}_n} \xrightarrow{d} W := S \cdot (B \circ \sigma^2)$ ,*
- (b) Part of Theorem 2 in (Ying, 1989): *On  $D[0, \tau]$  we have  $\sqrt{n}(\hat{S}_n - S) \xrightarrow{d} W = S \cdot (B \circ \sigma^2)$ .*

Denote by  $\hat{A}_n(t) = \sum_{i: X_{i:n} \leq t} \frac{\delta_{[i:n]}}{n-i+1}$  the Nelson-Aalen estimator for the cumulative hazard function  $A(t)$  and by  $\hat{G}_n$  the Kaplan-Meier estimator for the censoring survival function  $G$ . Note that  $\hat{H}_n = \hat{G}_n \hat{S}_n$  holds for the empirical survival function of  $H$  since, almost surely (a.s.), no survival time equals a censoring time:  $T_i \neq C_j$  a.s. for all  $i, j$ . The asymptotic covariance function  $\Gamma$  of  $W$  in Lemma 2.1 and a natural estimator  $\hat{\Gamma}_n$  are given by

$$\Gamma(u, v) = S(u) \left( \int_0^{u \wedge v} \frac{dA}{H_-} \right) S(v) \quad \text{and} \quad \hat{\Gamma}_n(u, v) = \hat{S}_n(u) \left( \int_0^{u \wedge v} \frac{d\hat{A}_n}{\hat{H}_{n-}} \right) \hat{S}_n(v).$$

The following lemma is helpful for an assessment of Condition (2.1) and for studentizations.

LEMMA 2.2. (a) *For all  $t \in [0, \tau]$  it holds that*

$$-\int_t^\tau \frac{d\hat{S}_n}{\hat{G}_{n-}} \xrightarrow{p} -\int_t^\tau \frac{dS}{G_-} \leq \infty.$$

(b) *In case of (2.1) we have*

$$\sup_{(u,v) \in [0,\tau]^2} |\hat{\Gamma}_n(u,v) - \Gamma(u,v)| \xrightarrow{p} 0.$$

### 3 Main Results

The limit distribution of the Kaplan-Meier process in Lemma 2.1 shall be assessed via bootstrapping. To this end, we independently draw  $n$  times with replacement from  $(X_1, \delta_1), \dots, (X_n, \delta_n)$  and denote the thus obtained bootstrap sample by  $(X_1^*, \delta_1^*), \dots, (X_n^*, \delta_n^*)$ . Throughout, denote by  $\Gamma_n^*, S_n^*$  etc. the obvious estimators but based on the bootstrap sample. Note that this requires a discontinuous extension of the above quantities. The following theorem is the basis of all later inference methods.

**Theorem 1.** *Let  $B$  denote a Brownian motion on  $[0, \tau]$  and suppose that (2.1) holds. Then we have, conditionally on  $X_1, X_2, \dots$ ,*

$$\sqrt{n}(S_n^* - \hat{S}_n) \xrightarrow{d} W = S \cdot (B \circ \sigma^2)$$

*on  $D[0, \tau]$  in probability.*

Many statistical applications involve a consistent variance estimator, e.g. Hall-Wellner or equal precision confidence bands for  $S$ ; cf. Andersen et al. (1993), p. 266. In order to asymptotically reproduce the same limit on the bootstrap side, the uniform consistency of a bootstrapped variance estimator (defined on the whole support  $[0, \tau]^2$  of the covariance function) needs to be verified. To this end, introduce the bootstrap version of  $\hat{\Gamma}_n$ , that is,

$$\Gamma_n^*(u, v) = S_n^*(u) \left( \int_0^{u \wedge v} \frac{dA_n^*}{H_{n-}^*} \right) S_n^*(v).$$

For all  $\varepsilon > 0$ , its uniform consistency (here and below always meaning conditional convergence in probability given  $X_1, X_2, \dots$  in probability) over all points  $(u, v) \in [0, \tau]^2 \setminus [\tau - \varepsilon, \tau]^2$  is an immediate consequence of Theorem 1 in combination with the continuous mapping theorem: Write the absolute value of the integral part minus its estimated counterpart as

$$\begin{aligned} & \left| \int_0^{u \wedge v} \frac{(\hat{H}_{n-} - H_{n-}^*) dA_n^* - H_{n-}^* d(\hat{A}_n - A_n^*)}{H_{n-}^* \hat{H}_{n-}} \right| \\ & \leq \frac{\sup_{(0, u \wedge v)} |\hat{H}_n - H_n^*|}{H_n^*((u \wedge v)-) \hat{H}_n((u \wedge v)-)} A_n^*(u \wedge v) + \left| \int_0^{u \wedge v} \frac{d(\hat{A}_n - A_n^*)}{\hat{H}_{n-}} \right|. \end{aligned}$$

The first term is asymptotically negligible due to Pòlya's theorem and the second term becomes small due to the continuous mapping theorem applied to the integral functional and the logarithm functional. Here the restriction to  $[0, \tau]^2 \setminus [\tau - \varepsilon, \tau]^2$  simplified the calculations since all denominators are asymptotically bounded away from zero.

For uniform consistency on the whole rectangle  $[0, \tau]^2$ , however, similar arguments as for the bootstrapped Kaplan-Meier process on  $[0, \tau]$  are required. Compared to (2.1), we postulate a slightly more restrictive censoring condition.

**LEMMA 3.1.** *Suppose that*

$$- \int_0^\tau \frac{dS}{G_-^3} < \infty. \quad (3.1)$$

*Then we have the following conditional uniform consistency given  $X_1, X_2, \dots$  in probability:*

$$\sup_{(u, v) \in [0, \tau]^2} |\Gamma_n^*(u, v) - \hat{\Gamma}_n(u, v)| \xrightarrow{p} 0 \quad \text{in probability as } n \rightarrow \infty. \quad (3.2)$$

**Remark 1.** *The proof of Lemma 3.1 shows that Condition (3.1) can be diminished to*

$$- \int_0^\tau \frac{dS}{G_-} - \int_0^\tau \frac{S^{1-\delta} dS}{G_-^{2+\delta}} < \infty$$

*for some  $\delta \in (0, 1)$ . This is due to the inequality  $(n \hat{H}_{n-}^3)^{-1} \leq (n^\delta \hat{H}_{n-}^{2+\delta})^{-1}$ .*

## 4 Applications

Applications of Theorem 1 concern confidence intervals for the mean residual life-time  $g(t) = \mathbb{E}[T - t \mid T > t]$  on compact sub-intervals  $[t_1, t_2] \subset [0, \tau]$  in case of  $\tau < \infty$  as well as confidence regions for the Lorenz curve  $L$  and the Gini index  $G$ . To this end, we apply the functional delta-method (e.g. Andersen et al., 1993, Theorem II.8.1) which in turn requires the Hadamard-differentiability of all involved statistical functionals.

### Confidence Bands for the Mean-Residual Lifetime Function

Let  $0 \leq t_1 \leq t_2$  and introduce the space  $C[t_1, \tau]$  of continuous functions on  $[t_1, \tau]$  equipped with the supremum norm as well as the subset

$$\tilde{C}[t_1, t_2] = \{f \in C[t_1, \tau] : \inf_{s \in [t_1, t_2]} |f(s)| > 0\} \subset C[t_1, \tau]$$

containing all continuous functions having a positive distance to the constant zero function on the interval  $[t_1, t_2]$ . Similarly, let

$$\tilde{D}[t_1, t_2] = \{f \in D[t_1, \tau] : \inf_{s \in [t_1, t_2]} |f(s)| > 0, \sup_{s \in [t_1, \tau]} |f(s)| < \infty\} \subset D[t_1, \tau]$$

be the extension of  $\tilde{C}[t_1, t_2]$  to possibly discontinuous, bounded càdlàg functions. For the notion of Hadamard-differentiability tangentially to subsets of  $D[t_1, \tau]$ , see Definition II.8.2, Theorem II.8.2 and Lemma II.8.3 in Andersen et al. (1993), p. 111f. The following lemma makes the functional delta-method available for applications to the mean residual life-time function.

LEMMA 4.1. *Let  $\tau < \infty$  and  $[t_1, t_2] \subset [0, \tau)$  be a compact interval. Then*

$$\psi : \tilde{D}[t_1, t_2] \rightarrow D[t_1, t_2], \quad \theta(\cdot) \mapsto \frac{1}{\theta(\cdot)} \int_{\cdot}^{\tau} \theta(s) ds$$

*is Hadamard-differentiable at each  $\theta \in \tilde{C}[t_1, t_2]$  tangentially to  $C^2[t_1, \tau]$  with continuous linear derivative  $d\psi(\theta) \cdot h \in D[t_1, t_2]$  given by*

$$(d\psi(\theta) \cdot h)(s) := \frac{1}{\theta(s)} \int_s^{\tau} h(u) du - h(s) \int_s^{\tau} \frac{\theta(u)}{\theta^2(s)} du.$$

As pointed out in Gill (1989) or Andersen et al. (1993), p. 110, the functional delta-method is established on the functional space  $D[t_1, \tau]$  (or subsets thereof) equipped with the supremum norm. However, in case of limiting processes with continuous sample paths, “weak convergence in the sense of the [Skorohod] metric and in the sense of the supremum norm are exactly equivalent” (Andersen et al., 1993). See also Problem 7 in Pollard (1984), p. 137. The convergence result of Theorem 1 combined with the functional  $\psi$  of Lemma 4.1 constitutes the following weak convergence.

LEMMA 4.2. *Suppose that (2.1) holds. On the Skorohod space  $D[t_1, t_2]$  we then have*

$$\sqrt{n} \left( \int_{\cdot}^{\tau} \frac{\hat{S}_n(u)}{\hat{S}_n(\cdot)} du - \int_{\cdot}^{\tau} \frac{S(u)}{S(\cdot)} du \right) \xrightarrow{d} U$$

*and, given  $X_1, X_2, \dots$ ,*

$$\sqrt{n} \left( \int_{\cdot}^{\tau} \frac{S_n^*(u)}{S_n^*(\cdot)} du - \int_{\cdot}^{\tau} \frac{\hat{S}_n(u)}{\hat{S}_n(\cdot)} du \right) \xrightarrow{d} U$$

in outer probability. The Gaussian process  $U$  has a.s. continuous sample paths, mean zero and covariance function

$$(r, s) \mapsto \int_{r \vee s}^{\tau} \int_{r \vee s}^{\tau} \frac{\Gamma(u, v)}{S(r)S(s)} du dv - \sigma^2(r \vee s)g(r)g(s),$$

where  $g(t) = \mathbb{E}[T_1 - t \mid T_1 > t] = \int_t^{\tau} \frac{S(u)}{S(t)} du$  is again the mean residual life-time function.

The previous lemma in combination with the continuous mapping theorem almost immediately gives rise to the construction of asymptotically valid confidence regions for the mean residual life-time function. According to the functional delta-method we may first apply, e.g. an arcsin- or log-transformation to ensure that only positive values are included in the confidence regions; cf. Section IV.1.3 in Andersen et al. (1993), p. 208ff. For ease of presentation, only the linear regions are stated below.

**Theorem 2.** Let  $0 \leq t_1 \leq t_2 < \tau$ . Choose any  $\alpha \in (0, 1)$  and suppose that (2.1) holds. An asymptotic two-sided  $(1 - \alpha)$ -confidence band for the mean residual life-time function  $(\mathbb{E}[T_1 - t \mid T_1 > t])_{t \in [t_1, t_2]}$  is given by

$$\left[ \int_{t_1}^{\tau} \frac{\hat{S}_n(u)}{\hat{S}_n(t_1)} du - \frac{q_{n_1, n_2}^{MRLT}}{\sqrt{n}}, \int_{t_2}^{\tau} \frac{\hat{S}_n(u)}{\hat{S}_n(t_2)} du + \frac{q_{n_1, n_2}^{MRLT}}{\sqrt{n}} \right]_{t \in [t_1, t_2]}$$

where  $q_{n_1, n_2}^{MRLT}$  is the  $(1 - \alpha)$ -quantile of the conditional law given  $(X_1, \delta_1), \dots, (X_n, \delta_n)$  of

$$\sqrt{n} \sup_{t \in [t_1, t_2]} \left| \int_t^{\tau} \frac{S_n^*(u)}{S_n^*(t)} du - \int_t^{\tau} \frac{\hat{S}_n(u)}{\hat{S}_n(t)} du \right|.$$

**Remark 2.** (a) Instead of using a transformation as indicated above Theorem 2, one could also employ a studentization using  $\hat{\Gamma}_n$  and  $\Gamma_n^*$ . Plugging these and consistent estimators for the other unknown quantities into the asymptotic variance representation yields consistent variance estimators for the statistic of interest. This yields a Gaussian process with asymptotic variance 1 at all points of time for the mean residual life-time estimates.

(b) In practice, the construction of confidence bands for the mean residual life-time function requires to choose  $t_2$  depending on the data: else, too large choices of  $t_2$  might result in  $\hat{S}_n(t_2) = 0$ , in which case the above estimator would not be well-defined.

## Confidence Regions for the Lorenz Curve and the Gini Index

Suppose  $S$  has compact support  $[0, \tau]$ , i.e. let this again be the smallest interval satisfying  $S(0) = 1$  and  $S(\tau) = 0$ . As estimators for the Lorenz curve and the Gini index we consider the plug-in estimates

$$\hat{L}_n(p) = \frac{1}{\hat{\mu}_n} \int_0^p (1 - \hat{S}_n(t))^{-1} dt \quad \text{and} \quad \hat{G}_n = \frac{\int_0^1 (u - \hat{L}_n(u)) du}{\int_0^1 u du},$$

where  $\hat{\mu}_n = \int_0^{\tau} s d\hat{S}_n(s)$ . The restricted and unscaled Lorenz curve estimator under independent right-censoring has been bootstrapped by Horvath and Yandell (1987). Tse (2006) discussed the large sample properties of the above Lorenz curve estimator (even under left-truncation) and also of the normalized estimated Gini index

$$\sqrt{n}(\hat{G}_n - G) = \sqrt{n} \left( \frac{\int_0^1 (u - \hat{L}_n(u)) du}{\int_0^1 u du} - \frac{\int_0^1 (u - L(u)) du}{\int_0^1 u du} \right) = 2\sqrt{n} \int_0^1 (L(u) - \hat{L}_n(u)) du.$$

Again equip all subsequent function spaces with the supremum norm. Let  $D_{\uparrow}[0, \tau] \subset D[0, \tau]$  be the set of all distribution functions on  $[0, \tau]$  with no atom in 0, and let  $D_{-}[0, \tau]$  be the set of all càglàd functions on  $[0, \tau]$ . First, we consider the normalized estimated Lorenz curve, i.e. the process  $W_n : \Omega \rightarrow [0, 1]$  given by

$$\begin{aligned} W_n(p) &= \sqrt{n} \left( \frac{1}{\hat{\mu}_n} \int_0^p (1 - \hat{S})^{-1}(s) ds - \frac{1}{\mu} \int_0^p (1 - S)^{-1}(s) ds \right) \\ &= \sqrt{n}(\hat{\mu}_n^{-1} \cdot (\Phi \circ \Psi \circ (1 - \hat{S}))(p) - \mu^{-1} \cdot (\Phi \circ \Psi \circ (1 - S))(p)). \end{aligned}$$

Here the functionals  $\Phi$  and  $\Psi$  are

$$\begin{aligned}\Phi : D_-[0, 1] &\mapsto C[0, 1], \quad h \mapsto \left(p \mapsto \int_0^p h(s) ds\right), \quad \text{and} \\ \Psi : D_+[0, \tau] &\mapsto D_-[0, 1], \quad k \mapsto k^{-1} \quad (\text{the left-continuous generalized inverse}).\end{aligned}$$

Suppose that  $S$  is continuously differentiable on its support with strictly positive derivative  $f$ , bounded away from zero. The Hadamard-differentiability of  $\Psi$  at  $(1 - S)$  tangentially to  $C[0, \tau]$  then holds according to Lemma 3.9.23 in van der Vaart and Wellner (1996), p. 386. Its derivative map is given by  $\alpha \mapsto -\frac{\alpha}{f} \circ (1 - S)^{-1}$ . The other functional  $\Phi$  is obviously Hadamard-differentiable at  $S^{-1} \in C[0, 1]$  tangentially to  $C[0, 1]$  since  $\Phi$  itself is linear and the domain of integration is bounded. Next,

$$\sqrt{n} \left( \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right) = \sqrt{n} (\Upsilon(\hat{g}_n(0)) - \Upsilon(g(0)))$$

where  $\Upsilon : (0, \infty) \rightarrow (0, \infty)$ ,  $r \mapsto \frac{1}{r}$ ,  $g(0) = \mathbb{E}[T - a \mid T > 0] = \mathbb{E}[T]$  is the mean-residual life-time function at 0 and  $\hat{g}_n(0)$  its estimated counterpart. Clearly,  $\Upsilon$  is (Hadamard-)differentiable and the required Hadamard-differentiability of  $(1 - S) \mapsto g(0)$  follows immediately from Lemma 4.1. Finally, the multiplication functional is also Hadamard-differentiable. All in all, we conclude that  $W_n = \sqrt{n}(\Xi(\hat{S}_n) - \Xi(S))$  for a functional  $\Xi : D[0, \tau] \rightarrow C[0, 1]$  which is Hadamard-differentiable at  $S$  tangentially to  $C[0, \tau]$ . Theorem 1 in combination with the functional  $\delta$ -method (for the bootstrap) immediately implies that  $W_n$  and  $W_n^*$  both converge in (conditional) distribution to the same continuous Gaussian process (in outer probability given  $\mathbf{X}$ ). Time-simultaneous inference procedures for the Lorenz curve, such as tests for equality and confidence bands are constructed straightforwardly.

Finally, the normalized estimated Gini index allows the representation

$$\sqrt{n}(\hat{G}_n - G) = 2\sqrt{n}(\{\Phi \circ \Xi\}(1) \circ \hat{S}_n - \{\Phi \circ \Xi\}(1) \circ S)$$

of which  $\{\Phi \circ \Xi\}(1)$  is again Hadamard-differentiable at  $S$  tangentially to  $C[0, \tau]$ . Hence, confidence intervals for  $G$  with bootstrap-based quantiles are constructed in the same way as before.

## 5 Discussion

In this article we established consistency of the bootstrap for Kaplan-Meier estimators on the whole support of the estimated survival function. By means of the functional delta-method this conditional weak convergence is transferred to Hadamard-differentiable functionals such as the mean-residual lifetime, the Lorenz curve or the Gini index. Further applications include the expected length of stay in the transient state (e.g. Grand and Putter 2015) or the probability of concordance (e.g. Pocock et al. 2012, Dobler and Pauly 2016).

This bootstrap consistency on the whole support may also be extended to more general inhomogeneous Markovian multistate models. Based on the martingale representation of Aalen-Johansen estimators for transition probability matrices (e.g. Andersen et al., 1993, p. 289), one could try to generalize the results of Gill (1983) to this setting. Here the notion of the ‘largest event times’ requires special attention as these may differ for different types of transitions. A reasonable first step towards such a generalization would be an analysis in competing risks set-ups where the support of each cumulative incidence function provides a natural domain to investigate weak convergences on. Once weak convergence of the estimators on the whole support is verified, martingale arguments similar to those of Akritas (1986) and Gill (1983) may be employed in order to obtain such (now conditional) weak convergences for the resampled Aalen-Johansen estimator using a variant of Efron’s bootstrap. In more general Markovian multi-state models we could independently draw with replacement from the sample that contains all individual *trajectories* rather than single observed transitions in order to not corrupt the dependencies within each individual; see for example Tattar and Vaman (2012) for a similar suggestion. Applications of this theory could include inference on more refined variants of the probability of concordance or the expected length of stay. Considering a progressive disease in a two-sample situation, for instance, we would like to compare the probability that

an individual of group one remains longer in a less severe disease state than an individual of group two. Accurate inference procedures for the mean residual life-time in a state of disability given any state at present time offers another kind of application.

## Acknowledgements

The author would like to thank Markus Pauly (University of Ulm) for helpful discussions.

## Appendix

Some of the following proofs (Appendix A) rely on the ideas of Gill (1983). In order to also apply (variants of) his lemmata in our bootstrap context, Appendix B below contains all required results. ‘Tightness’ in the support’s right boundary  $\tau$  for the bootstrapped Kaplan-Meier estimator is essentially shown via a bootstrap version of the approximation theorem for truncated estimators as in Theorem 3.2 in Billingsley (1999); cf. Appendix C. Define by  $Y(u) = n\hat{H}_{n-}(u)$  the process counting the number of individuals at risk of dying, and by  $Y^*(u)$  its bootstrap version.

## A Proofs

*Proof of Lemma 2.2.* Proof of (a): Let  $t < \tau$  and suppose (2.1) holds. By the continuous mapping theorem and the boundedness away from zero of  $\frac{1}{G}$  on  $[0, t]$ , it clearly follows that  $-\int_0^t \frac{d\hat{S}_n}{\hat{G}_{n-}} \xrightarrow{p} -\int_0^t \frac{dS}{G_-}$  as  $n \rightarrow \infty$ . Letting  $t \uparrow \tau$ , the right-hand side converges towards  $-\int_0^\tau \frac{dS}{G_-} < \infty$ . It remains to apply Theorem 3.2 of Billingsley (1999) in order to verify the assertion for  $t = 0$  and hence for all  $t \leq \tau$  by the continuous mapping theorem. Thus, we show that for all  $\varepsilon > 0$ ,

$$\lim_{t \uparrow \tau} \limsup_{n \rightarrow \infty} P\left(-\int_t^\tau \frac{d\hat{S}_n}{\hat{G}_{n-}} > \varepsilon\right) = 0.$$

Let  $\hat{T}_n$  again be the largest observation among  $X_1, \dots, X_n$  and define, for any  $\beta > 0$ ,

$$B_\beta := \{\hat{S}_n(s) \leq \beta^{-1}S(t) \text{ and } \hat{H}_n(s-) \geq \beta H(s-) \text{ for all } s \in [0, \hat{T}_n]\}.$$

By Lemmata B.1 and B.2, the probability  $p_\beta := 1 - P(B_\beta) \leq \beta + \frac{e}{\beta} \exp(-1/\beta)$  is arbitrary small for sufficiently small  $\beta > 0$ . Hence, by Theorem 1.1 of Stute and Wang (1993) (applied for the concluding convergence),

$$\begin{aligned} P\left(-\int_t^\tau \frac{d\hat{S}_n}{\hat{G}_{n-}} > \varepsilon\right) &= P\left(-\int_t^\tau \frac{\hat{S}_n - d\hat{S}_n}{\hat{H}_{n-}} > \varepsilon\right) \\ &\leq P\left(-\beta^{-2} \int_t^\tau \frac{S_- - d\hat{S}_n}{H_-} > \varepsilon\right) + p_\beta \\ &= P\left(-\beta^{-2} \int_t^\tau \frac{d\hat{S}_n}{G_-} > \varepsilon\right) + p_\beta \rightarrow P\left(-\beta^{-2} \int_t^\tau \frac{dS}{G_-} > \varepsilon\right) + p_\beta. \end{aligned}$$

For large  $t < \tau$  and by the continuity of  $S$ , the far right-hand side of the previous display equals  $p_\beta$ .

Proof of (b): First note that the uniform convergences in probability in Theorems IV.3.1 and IV.3.2 of Andersen et al. (1993), p. 261ff., yield, for any  $\varepsilon > 0$ ,

$$\sup_{(u,v) \in [0, \tau - \varepsilon]^2} |\hat{\Gamma}_n(u, v) - \Gamma(u, v)| \xrightarrow{p} 0 \quad \text{as } n \rightarrow \infty.$$



Further, the dominated convergence theorem and  $S_- dA = -dS$  show that

$$\Gamma(u, v) = - \int_0^\tau \mathbf{1}\{w \leq u \wedge v\} \frac{S(u)S(v)}{S(w-)S(w-)} \frac{dS(w)}{G(w-)} \rightarrow - \int_0^\tau 0 \frac{dS}{G_-} = 0$$

as  $u, v \rightarrow \tau$ . Hence, it remains to verify the remaining condition (3.8) of Theorem 3.2 in Billingsley (1999) in order to conclude this proof. That is, for each positive  $\delta$  we show

$$\lim_{u, v \rightarrow \tau} \limsup_{n \rightarrow \infty} P\left(\sup_{(u, v) \in [0, \tau]^2} |\hat{\Gamma}_n(u, v) - \hat{\Gamma}_n(\tau, \tau)| \geq \delta\right) = 0.$$

To this end, rewrite  $\hat{\Gamma}_n(u, v) - \hat{\Gamma}_n(\tau, \tau)$  as

$$\int_{u \wedge v}^\tau \frac{\hat{S}_n(\tau)\hat{S}_n(\tau)}{\hat{S}_n(w-)\hat{S}_n(w-)} \frac{d\hat{S}_n(w)}{\hat{G}_n(w-)} + \int_0^{u \wedge v} \frac{d\hat{S}_n}{\hat{S}_n^2 \hat{G}_n} (\hat{S}_n^2(\tau) - \hat{S}_n(u)\hat{S}_n(v)).$$

The left-hand integral is bounded in absolute value by  $-\int_{u \wedge v}^\tau \frac{d\hat{S}_n}{\hat{G}_n}$  which goes to  $-\int_{u \wedge v}^\tau \frac{dS}{G_-}$  in probability as  $n \rightarrow \infty$  by (a). For large  $u, v$  this is arbitrarily small.

The remaining integral is bounded in absolute value by

$$-\int_0^{u \wedge v} \frac{\hat{S}_n(u)\hat{S}_n(v)}{\hat{S}_n^2 \hat{G}_n} \frac{d\hat{S}_n}{\hat{G}_n} = -n^2 \int_0^{u \wedge v} \frac{\hat{S}_n(u)\hat{S}_n(v)}{Y^2} \hat{G}_n d\hat{S}_n.$$

By Lemmata B.1 and B.2 this integral is bounded from above by

$$-\int_0^{u \wedge v} \frac{S(\hat{T}_n \wedge u)S(\hat{T}_n \wedge v)}{H_-^2} G_- d\hat{S}_n = -\int_0^{u \wedge v} \frac{S(\hat{T}_n \wedge u)S(\hat{T}_n \wedge v)}{S_-^2 G_-} d\hat{S}_n$$

on a set with arbitrarily high probability. For sufficiently large  $n$  we also have  $\hat{T}_n > u \wedge v$  with arbitrarily high probability. Next, Theorem 1.1 in Stute and Wang (1993) yields

$$-\int_0^{u \wedge v} \frac{S(u)S(v)}{S_-^2 G_-} d\hat{S}_n \xrightarrow{p} -\int_0^{u \wedge v} \frac{S(u)S(v)}{S_-^2 G_-} dS \quad \text{as } n \rightarrow \infty.$$

As above the dominated convergence theorem shows the negligibility of this integral as  $u, v \rightarrow \tau$ .  $\square$

*Proof of Theorem 1.* For the proof of weak convergence of the bootstrapped Kaplan-Meier estimator on each Skorohod space  $D[0, t]$ ,  $t < \tau$ , see e.g. Akritas (1986), Lo and Singh (1986) or Horvath and Yandell (1987). By defining these processes as constant functions after  $t$ , the convergences equivalently hold on  $D[0, \tau]$ . This takes care of Condition (a) in Lemma C.1, while (c) is obviously fulfilled by the continuity of the limit Gaussian process.

To close the indicated gap for the bootstrapped Kaplan-Meier process on the whole support  $[0, \tau]$ , it remains to analyze Condition (b). This is first verified for the truncated process by following the strategy of Gill (1983) while applying the martingale theory of Akritas (1986) for the bootstrapped counting processes. Thus, the truncation technique of Lemma C.1 shows the convergence in distribution of the truncated process. Finally, the negligibility of the remainder term is shown similarly as in Ying (1989).

We will make use of the fact that our martingales, stopped at arbitrary stopping times, retain the martingale property; cf. Andersen et al. (1993), p. 70, for sufficient conditions on this matter. Similarly to the largest event or censoring time  $\hat{T}_n$ , introduce the largest bootstrap time  $T_n^* = \max_{i=1, \dots, n} X_i^*$ , being an integrable stopping time with respect to the filtration of Akritas (1986) who used Theorem 3.1.1 of Gill (1980): Hence, we choose the filtration given by

$$\mathcal{F}_t := \{X_i, \delta_i, \delta_i^* \mathbf{1}\{X_i^* \leq t\}, X_i^* \mathbf{1}\{X_i^* \leq t\} : i = 1, \dots, n\}, \quad 0 \leq t \leq \tau;$$

see also Gill (1980), p. 26, for a similar minimal filtration. Note that we did not include the indicators  $\mathbf{1}\{X_i^* \leq t\}$  into the filtration since their values are already determined by all the  $X_i^* \mathbf{1}\{X_i^* \leq t\}$ : According to our assumptions,  $X_i^* > 0$  a.s. for all  $i = 1, \dots, n$ .

We would first like to verify condition (b) in Lemma C.1 for the stopped bootstrap Kaplan-Meier process. That is, for each  $\varepsilon > 0$  and an arbitrary subsequence  $(n') \subset (n)$  there is another subsequence  $(n'') \subset (n')$  such that

$$\begin{aligned} & \lim_{t \uparrow \tau} \limsup_{n'' \rightarrow \infty} P\left(\sup_{t \leq s < T_n^*} \sqrt{n''} |(S_n^* - \hat{S}_n)(s) - (S_n^* - \hat{S}_n)(t)| > \varepsilon \mid \mathbf{X}\right) \\ & \leq \lim_{t \uparrow \tau} \limsup_{n'' \rightarrow \infty} P\left(\sup_{t \leq s < \hat{T}_n} \sqrt{n''} |(S_n^* - \hat{S}_n)(s \wedge T_n^*) - (S_n^* - \hat{S}_n)(t \wedge T_n^*)| > \varepsilon \mid \mathbf{X}\right) \\ & = 0 \quad \text{a.s.} \end{aligned} \tag{A.1}$$

for all  $\varepsilon > 0$ . Here  $\sigma(\mathbf{X}) = \mathcal{F}_0$  summarizes the collected data. Due to the boundedness away from zero, i.e.  $\inf_{t \leq s < \hat{T}_n} \hat{S}_n(s) > 0$ , we may rewrite the bootstrap process  $\sqrt{n}(S_n^* - \hat{S}_n)(s) = \sqrt{n}\left(\frac{S_n^*(s)}{\hat{S}_n(s)} - 1\right)\hat{S}_n(s)$  for each  $s \in [t, \hat{T}_n)$  of which the bracket term is a square integrable martingale; see Akritas (1986) again. Hence, the term  $\sqrt{n}(S_n^* - \hat{S}_n)(s)$  in (A.1) equals

$$M_n^*(s)\hat{S}_n(s \wedge T_n^*) := \sqrt{n}\left(\frac{S_n^*(s \wedge T_n^*)}{\hat{S}_n(s \wedge T_n^*)} - 1\right)\hat{S}_n(s \wedge T_n^*), \tag{A.2}$$

whereof  $(M_n^*(s))_{s \in [0, \hat{T}_n)}$  is again a square integrable martingale. Indeed, its predictable variation process evaluated at the stopping time  $s = T_n^*$  is finite (having the sufficient condition of Andersen et al. (1993), p. 70, for a stopped martingale to be a square integrable martingale in mind): The predictable variation is given by

$$s \mapsto \langle M_n^* \rangle(s) = \int_0^{s \wedge T_n^*} \left(\frac{S_{n-}^*}{\hat{S}_n}\right)^2 \frac{(1 - \Delta \hat{A}_n) d\hat{A}_n}{H_{n-}^*},$$

where  $H_n^*$  is the empirical survival function of  $X_1^*, \dots, X_n^*$  and  $\Delta f$  denotes the increment process  $s \mapsto f(s+) - f(s-)$  of a monotone function  $f$ . The supremum in (A.1) is bounded by

$$\sup_{t \leq s < \hat{T}_n} |M_n^*(s) - M_n^*(t)|\hat{S}_n(s \wedge T_n^*) + \sup_{t \leq s < \hat{T}_n} |M_n^*(t)||\hat{S}_n(s \wedge T_n^*) - \hat{S}_n(t \wedge T_n^*)|$$

of which the right-hand term is not greater than  $|M_n^*(t)|\hat{S}_n(t \wedge T_n^*)$ . By the convergence in distribution of the bootstrapped Kaplan-Meier estimator on each  $D[0, \tilde{\tau}]$ ,  $\tilde{\tau} < \tau$ , we have convergence in conditional distribution of  $M_n^*(t)\hat{S}_n(t \wedge T_n^*)$  given  $\mathbf{X}$  towards  $N(0, S^2(t)\Gamma(t, t))$  in probability. Hence,

$$\lim_{n'' \rightarrow \infty} P(|M_n^*(t)\hat{S}_n(t \wedge T_n^*)| > \varepsilon/2 \mid \mathbf{X}) \rightarrow 1 - N(0, \Gamma(t, t))(-\varepsilon/2, \varepsilon/2)$$

almost surely along subsequences  $(n'')$  of arbitrary subsequences  $(n') \subset (n)$ . Since the variance of the normal distribution in the previous display goes to zero as  $t \uparrow \tau$ , cf. (2.4) in Gill (1983), the above probability vanishes as  $t \uparrow \tau$ .

By Lemma B.3, the remainder  $\sup_{t \leq s < \hat{T}_n} |M_n^*(s) - M_n^*(t)|\hat{S}_n(s \wedge T_n^*)$  is not greater than

$$2 \sup_{t \leq s < \hat{T}_n} \left| \int_t^s \hat{S}_n(u) dM_n^*(u) \right|. \tag{A.3}$$

Since, given  $\mathbf{X}$ ,  $\hat{S}_n$  is a bounded and predictable process, this integral is a square integrable martingale on  $[t, \hat{T}_n)$ . We proceed as in Gill (1983) by applying Lenglart's inequality, cf. Section II.5.2 in Andersen et al. (1993): For

each  $\eta > 0$  we have

$$\begin{aligned} P\left(\sup_{t \leq s < T_n^* \wedge \tau} \left| \int_t^s \hat{S}_n dM_n^* \right| > \varepsilon \mid \mathbf{X}\right) \\ \leq \frac{\eta}{\varepsilon^2} + P\left(\left| \int_t^{\tau \wedge T_n^*} S_{n-}^{*2} \frac{(1 - \Delta \hat{A}_n) d\hat{A}_n}{\hat{H}_{n-}^*} \right| > \eta \mid \mathbf{X}\right). \end{aligned} \quad (\text{A.4})$$

We intersect the event on the right-hand side of (A.4) with  $B_{H,n,\beta}^* := \{H_n^*(s-) \geq \beta \hat{H}_n(s-) \text{ for all } s \in [t, T_n^*]\}$  and also with  $B_{S,n,\beta}^* := \{S_n^*(s) \leq \beta^{-1} \hat{S}_n(s) \text{ for all } s \in [t, T_n^*]\}$ . According to Lemmata B.1 and B.2, the conditional probabilities of these events are at least  $1 - \exp(1 - 1/\beta)/\beta$  and  $1 - \beta$ , respectively, for any  $\beta \in (0, 1)$ . Thus, (A.4) is less than or equal to

$$\frac{\eta}{\varepsilon^2} + \beta + \frac{\exp(1 - 1/\beta)}{\beta} + \mathbf{1}\left\{\beta^{-3} \left| \int_t^{\tau \wedge \hat{T}_n} \hat{S}_{n-}^2 \frac{(1 - \Delta \hat{A}_n) d\hat{A}_n}{\hat{H}_{n-}} \right| > \eta\right\}. \quad (\text{A.5})$$

In order to show the almost sure negligibility of the indicator function as  $n \rightarrow \infty$  and then  $t \uparrow \tau$ , we analyze the corresponding convergence of the integral. Since  $-d\hat{S}_n = \hat{S}_{n-} d\hat{A}_n$ , the integral is less than or equal to

$$- \int_t^\tau \frac{\hat{S}_{n-} d\hat{S}_n}{\hat{H}_{n-}} = - \int_t^\tau \frac{d\hat{S}_{n-}}{\hat{G}_{n-}}.$$

Lemma 2.2 implies that for each subsequence  $(n') \subset (n)$  there is another subsequence  $(n'') \subset (n')$  such that  $-\int_t^\tau \frac{d\hat{S}_{n-}}{\hat{G}_{n-}} \rightarrow -\int_t^\tau \frac{dS}{G_-}$  a.s. for all  $t \in [0, \tau] \cap \mathbb{Q}$  along  $(n'')$ . Due to  $P(Z_1 \in \mathbb{Q}) = 0$ , the same convergence holds for all  $t \leq \tau$ . Letting now  $t \uparrow \tau$  shows that the indicator function in (A.5) vanishes almost surely in limit superior along  $(n'')$  if finally  $t \uparrow \tau$ . The remaining terms are arbitrarily small for sufficiently small  $\eta, \beta > 0$ . Hence, all conditions of Lemma C.1 are met and the assertion follows for the stopped process

$$(\mathbf{1}\{s < T_n^*\} \sqrt{n}(S_n^*(s) - \hat{S}_n(s)) + \mathbf{1}\{s \geq T_n^*\} \sqrt{n}(S_n^*(T_n^* -) - \hat{S}_n(T_n^* -)))_{s \in [0, \tau]}.$$

Finally, we show the asymptotic negligibility of

$$\begin{aligned} \sup_{T_n^* \leq s \leq \tau} \sqrt{n}|S_n^*(s) - \hat{S}_n(s)| &\leq \sup_{T_n^* \leq s \leq \tau} \sqrt{n}(S_n^*(s) + \hat{S}_n(s)) \\ &= \sqrt{n}S_n^*(T_n^*) + \sqrt{n}\hat{S}_n(T_n^*); \end{aligned}$$

cf. Ying (1989) for similar considerations. Again by Lemma B.1, we have for any  $\varepsilon > 0, \beta \in (0, 1)$  that

$$\begin{aligned} P(\sqrt{n}S_n^*(T_n^*) + \sqrt{n}\hat{S}_n(T_n^*) > \varepsilon \mid \mathbf{X}) \\ \leq P(\sqrt{n}S_n^*(T_n^*) > \varepsilon/2 \mid \mathbf{X}) + P(\sqrt{n}\hat{S}_n(T_n^*) > \varepsilon/2 \mid \mathbf{X}) \\ \leq P(\sqrt{n}\hat{S}_n(T_n^*) > \beta\varepsilon/2 \mid \mathbf{X}) + P(\sqrt{n}\hat{S}_n(T_n^*) > \varepsilon/2 \mid \mathbf{X}) + \beta. \end{aligned}$$

Define the generalized inverse  $\hat{S}_n^{-1}(u) := \inf\{s \leq \tau : \hat{S}_n(s) \geq u\}$ . The independence of the bootstrap drawings as well as arguments of quantile transformations yield

$$\begin{aligned} P(\sqrt{n}\hat{S}_n(T_n^*) > \varepsilon \mid \mathbf{X}) &= P(X_1^* < \hat{S}_n^{-1}(\varepsilon/\sqrt{n}) \mid \mathbf{X})^n \\ &= \left[1 - \frac{1}{n} |\{i : X_i \geq \hat{S}_n^{-1}(\varepsilon/\sqrt{n})\}| \right]^n. \end{aligned}$$

The cardinality in the display goes to infinity in probability, and hence almost surely along subsequences. Indeed, for any constant  $C > 0$ ,

$$\begin{aligned}
& P(|\{i : X_i \geq \hat{S}_n^{-1}(\varepsilon/\sqrt{n})\}| \geq C) \\
&= P(|\{i : \hat{S}_n(X_i) \geq \varepsilon/\sqrt{n}\}| \geq C) \\
&\geq P(|\{i : \hat{H}_n(X_i) \geq \varepsilon/\sqrt{n}\}| \geq C) \\
&= P\left(\left|\left\{i : \frac{i-1}{n} \geq \frac{\varepsilon}{\sqrt{n}}\right\}\right| \geq C\right) \\
&= \mathbf{1}\left\{\left|\left\{i : \frac{i-1}{n} \geq \frac{\varepsilon}{\sqrt{n}}\right\}\right| \geq C\right\} = \mathbf{1}\{|\{[\varepsilon\sqrt{n}] + 1, \dots, n\}| \geq C\}.
\end{aligned}$$

Clearly, this indicator function goes to 1 as  $n \rightarrow \infty$ . □

*Proof of Lemma 3.1.* For the most part, we follow the lines of the above proof of Lemma 2.2 by verifying condition (3.8) of Theorem 3.2 in Billingsley (1999). To point out the major difference to the previous proof, we consider

$$-\int_{u \wedge v}^{\tau} \frac{dS_n^*}{G_{n-}^*} = \int_{u \wedge v}^{\tau} \frac{S_{n-}^* dA_n^*}{G_{n-}^*} = \int_{u \wedge v}^{\tau} \frac{S_{n-}^*}{G_{n-}^*} J^* d(A_n^* - \hat{A}_n) + \int_{u \wedge v}^{\tau} \frac{S_{n-}^*}{G_{n-}^*} J^* d\hat{A}_n,$$

where  $J^*(u) = \mathbf{1}\{Y^*(u) > 0\}$ . The arguments of Akritas (1986) show that  $\int_{u \wedge v}^{\tau} \frac{S_{n-}^*}{G_{n-}^*} J^* d(A_n^* - \hat{A}_n)$  is a square-integrable martingale with predictable variation process given by

$$t \mapsto \int_{u \wedge v}^t \frac{S_{n-}^{*2}}{G_{n-}^{*2}} \frac{J^*}{Y^*} (1 - \Delta \hat{A}_n) d\hat{A}_n.$$

After writing  $S_n^* G_n^* = H_n^*$ , a two-fold application of Lemmata B.1 and B.2 (at first to the bootstrap quantities  $S_n^*$  and  $H_n^*$ , then to the Kaplan-Meier estimators  $\hat{S}_n$  and  $\hat{H}_n$ ) show that the predictable variation in the previous display is bounded from above by

$$-\beta^{-13} \frac{1}{n} \int_{u \wedge v}^t \frac{S_-^3}{H_-^3} d\hat{S}_n = -\beta^{-13} \frac{1}{n} \int_{u \wedge v}^t \frac{d\hat{S}_n}{G_-^3}$$

on a set with arbitrarily large probability depending on  $\beta \in (0, 1)$ . Here we also used that  $\hat{S}_{n-} d\hat{A}_n = d\hat{S}_n$ . Due to (3.1), Theorem 1.1 of Stute and Wang (1993) yields

$$-\int_{u \wedge v}^t \frac{d\hat{S}_n}{G_-^3} \xrightarrow{p} -\int_{u \wedge v}^t \frac{dS}{G_-^3} < \infty$$

and hence the asymptotic negligibility of the predictable variation process in probability. By Rebolledo's theorem (Theorem II.5.1 in Andersen et al., 1993, p. 83),  $\int_{u \wedge v}^{\tau} \frac{S_{n-}^*}{G_{n-}^*} J^* d(A_n^* - \hat{A}_n)$  hence goes to zero in conditional probability. The remaining integral  $\int_{u \wedge v}^{\tau} \frac{S_{n-}^*}{G_{n-}^*} J^* d\hat{A}_n$  is treated similarly with Lemmata B.1 and B.2 and Theorem 1.1 of Stute and Wang (1993) yielding a bound in terms of  $\int_{u \wedge v}^{\tau} \frac{dS}{G_-}$ . This is arbitrarily small for sufficiently large  $u, v < \tau$ . □

*Proof of Lemma 4.1.* Proof of (b): Throughout, the functional spaces  $D[t_1, \tau]$  and  $\tilde{D}[t_1, t_2]$  are equipped with the supremum norm. For some sequences  $t_n \downarrow 0$  and  $h_n \rightarrow h$  in  $D[t_1, \tau]$  such that  $\theta + t_n h_n \in \tilde{D}[t_1, t_2]$ , consider the

supremum distance

$$\sup_{s \in [t_1, t_2]} \left| \frac{1}{t_n} [\psi(\theta + t_n h_n)(s) - \psi(\theta)(s)] - (d\psi(\theta) \cdot h)(s) \right|. \quad (\text{A.6})$$

The proof is concluded if (A.6) goes to zero. For an easier access the expression in the previous display is first analyzed for each fixed  $s \in [t_1, t_2]$ :

$$\begin{aligned} & \frac{1}{t_n} [\psi(\theta + t_n h_n)(s) - \psi(\theta)(s)] - (d\psi(\theta) \cdot h)(s) \\ &= \frac{1}{t_n} \frac{1}{\theta(s) + t_n h_n(s)} \frac{1}{\theta(s)} \\ & \quad \times \left[ \theta(s) \int_s^\tau (\theta(u) + t_n h_n(u)) du - (\theta(s) + t_n h_n(s)) \int_s^\tau \theta(u) du \right] \\ & \quad - \frac{1}{\theta(s)} \int_s^\tau h(u) du + h(s) \int_s^\tau \frac{\theta(u)}{\theta^2(s)} du \\ &= \frac{1}{\theta(s) + t_n h_n(s)} \int_s^\tau h_n(u) du - \frac{h_n(s)}{\theta(s) + t_n h_n(s)} \frac{1}{\theta(s)} \int_s^\tau \theta(u) du \\ & \quad - \frac{1}{\theta(s)} \int_s^\tau h_n(u) du + h_n(s) \int_s^\tau \frac{\theta(u)}{\theta^2(s)} du \\ & \quad - \frac{1}{\theta(s)} \int_s^\tau (h(u) - h_n(u)) du - (h(s) - h_n(s)) \int_s^\tau \frac{\theta(u)}{\theta^2(s)} du \\ &= - \int_s^\tau h_n(u) du \frac{t_n h_n(s)}{[\theta(s) + t_n h_n(s)] \theta(s)} + h_n(s) \int_s^\tau \frac{\theta(u)}{\theta^2(s)} du \frac{t_n h_n(s)}{\theta(s) + t_n h_n(s)} \\ & \quad - \frac{1}{\theta(s)} \int_s^\tau (h(u) - h_n(u)) du - (h(s) - h_n(s)) \int_s^\tau \frac{\theta(u)}{\theta^2(s)} du. \end{aligned} \quad (\text{A.7})$$

For large  $n$ , each denominator is bounded away from zero: To see this, denote  $\varepsilon := \inf_{s \in [t_1, t_2]} |\theta(s)|$  and  $C := \sup_{s \in [t_1, t_2]} |h(u)|$ . Thus,

$$\sup_{s \in [t_1, t_2]} |h_n(s)| \leq \sup_{s \in [t_1, t_2]} |h_n(s) - h(u)| + \sup_{s \in [t_1, t_2]} |h(s)| \leq \varepsilon + C$$

for each  $n$  large enough. It follows that, for each such  $n$  additionally satisfying  $t_n \leq \varepsilon(2\varepsilon + 2C)^{-1}$ , the denominators are bounded away from zero, in particular,  $\inf_{s \in [t_1, t_2]} |\theta(s) + t_n h_n(s)| \geq \varepsilon/2$ . Thus, taking the suprema over  $s \in [t_1, t_2]$ , the first two terms in (A.7) become arbitrarily small by letting  $t_n$  be sufficiently small. The remaining two terms converge to zero since  $\sup_{s \in [t_1, t_2]} |h_n(s) - h(s)| \rightarrow 0$  and  $\sup_{u \in [t_1, \tau]} |\theta(u)| < \infty$ . Note here that

$$\int_{t_1}^\tau |h(u) - h_n(u)| du \leq \sup_{u \in [t_1, \tau]} |h(u) - h_n(u)| (\tau - t_1) \rightarrow 0$$

due to  $\tau < \infty$ . □

*Proof of Lemma 4.2.* The convergences are immediate consequences of the functional delta-method, Theorem 1 and the bootstrap version of the delta-method; cf. Section 3.9 in van der Vaart and Wellner (1996). Simply note that all considered survival functions are elements of  $D_{<\infty} \cap \tilde{D}[t_1, t_2]$  (on increasing sets with probability tending to one) and that the survival function of the life-times is assumed continuous and bounded away from zero on compact subsets of  $[0, \tau)$ . Further, there is a version of the limit Gaussian processes with almost surely continuous sample paths.

For the representation of the variance of the limit distribution in part (a) we refer to van der Vaart and Wellner (1996), p. 383 and 397. The asymptotic covariance structure in part (b) is easily calculated using Fubini's theorem – for its applicability note that the variances  $\Gamma(r, r)$  of the limit process  $W$  of the Kaplan-Meier estimator exist at all points of time  $r \in [0, \tau]$ . Thus, since  $W$  is a zero-mean process, we have for any  $0 \leq r \leq s < \tau$ ,

$$\begin{aligned} & \text{cov} \left( \int_r^\tau \frac{W(u)}{S(r)} du - \int_r^\tau \frac{W(r)S(u)}{S^2(r)} du, \int_s^\tau \frac{W(v)}{S(s)} dv - \int_s^\tau \frac{W(s)S(v)}{S^2(s)} dv \right) \\ &= \int_r^\tau \int_s^\tau \left[ \Gamma(u, v) - \frac{S(u)}{S(r)} \Gamma(r, v) - \frac{S(v)}{S(s)} \Gamma(s, u) + \frac{S(u)S(v)}{S(r)S(s)} \Gamma(r, s) \right] \frac{dudv}{S(r)S(s)}. \end{aligned}$$

Inserting the definition  $\Gamma(r, s) = S(r)S(s)\sigma^2(r \wedge s)$  and splitting the first integral into  $\int_r^\tau = \int_r^s + \int_s^\tau$  yields that the last display equals

$$\begin{aligned} & \int_r^\tau \int_s^\tau \frac{S(u)S(v)}{S(r)S(s)} [\sigma^2(u \wedge v) - \sigma^2(r \wedge v) - \sigma^2(s \wedge u) + \sigma^2(r \wedge s)] dudv \\ &= \int_s^\tau \int_s^\tau \frac{S(u)S(v)}{S(r)S(s)} [\sigma^2(u \wedge v) - \sigma^2(r) - \sigma^2(s) + \sigma^2(r)] dudv \\ & \quad + \int_r^s \int_s^\tau \frac{S(u)S(v)}{S(r)S(s)} [\sigma^2(u) - \sigma^2(r) - \sigma^2(u) + \sigma^2(r)] dudv \\ &= \int_s^\tau \int_s^\tau \frac{\Gamma(u, v)}{S(r)S(s)} dudv - \sigma^2(r \vee s)g(r)g(s). \end{aligned}$$

□

*Proof of Theorem 2.* The theorem follows from Lemma 4.2 combined with the continuous mapping theorem applied to the supremum functional  $D[t_1, t_2] \rightarrow \mathbb{R}, f \mapsto \sup_{t \in [t_1, t_2]} |f(t)|$  which is continuous on  $C[t_1, t_2]$ . For the connection between the consistency of a bootstrap distribution of a real statistic and the consistency of the corresponding tests (and the equivalent formulation in terms of confidence regions), see Lemma 1 in Janssen and Pauls (2003). □

## B Adaptations of Gill's (1983) Lemmata

Abbreviate again the sigma algebra containing all the information of the original sample as  $\mathbf{X} := \sigma(X_i, \delta_i : i = 1, \dots, n)$ . The proofs in Appendix A rely on bootstrap versions of Lemmata 2.6, 2.7 and 2.9 in Gill (1983). Since those are stated under the assumption of a continuous distribution function  $S$ , but ties in the bootstrap sample are inevitable, these lemmata need a slight extension. For completeness, parts (a) of the following two Lemmata correspond to the original Lemmata 2.6 and 2.7 in Gill (1983).

LEMMA B.1 (Extension of Lemma 2.6 in Gill, 1983). *For any  $\beta \in (0, 1)$ ,*

- (a)  $P(\hat{S}_n(t) \leq \beta^{-1}S(t) \text{ for all } t \leq \hat{T}_n) \geq 1 - \beta,$
- (b)  $P(S_n^*(t) \leq \beta^{-1}\hat{S}_n(t) \text{ for all } t \leq T_n^* \mid \mathbf{X}) \geq 1 - \beta \text{ almost surely.}$

*Proof of (b).* All equalities and inequalities concerning conditional expectations are understood as to hold almost surely. As in the proof of Theorem 1,  $(S_n^*(t \wedge T_n^*)/\hat{S}_n(t \wedge T_n^*))_{t \in [0, \hat{T}_n]}$  defines a right-continuous martingale

for each fixed  $n$  and for almost every given sample  $\mathbf{X}$ . Hence, Doob's  $L_1$ -inequality (e.g. Revuz and Yor, 1999, Theorem 1.7 in Chapter II) yields for each  $\beta \in (0, 1)$

$$\begin{aligned} P(\sup_{t \in [0, \hat{T}_n)} S_n^*(t \wedge T_n^*) / \hat{S}_n(t \wedge T_n^*) \geq \beta^{-1} \mid \mathbf{X}) \\ \leq \beta \sup_{t \in [0, \hat{T}_n)} E(S_n^*(t \wedge T_n^*) / \hat{S}_n(t \wedge T_n^*) \mid \mathbf{X}) \\ = \beta E(S_n^*(0) / \hat{S}_n(0) \mid \mathbf{X}) = \beta. \end{aligned}$$

This implies  $P(S_n^* \leq \beta^{-1} \hat{S}_n \text{ on } [0, T_n^*) \mid \mathbf{X}) \geq 1 - \beta$ . It remains to extend this result to the interval's endpoint. If the observation corresponding to  $T_n^*$  is uncensored, we have  $0 = S_n^*(T_n^*) \leq \beta^{-1} \hat{S}_n(T_n^*)$ . Else, the event of interest  $\{S_n^* \leq \beta^{-1} \hat{S}_n \text{ on } [0, T_n^*)\}$  (given  $\mathbf{X}$ ) implies that

$$S_n^*(T_n^*) = S_n^*(T_n^* -) \leq \beta^{-1} \hat{S}_n(T_n^* -) = \beta^{-1} \hat{S}_n(T_n^*).$$

Thus, for given  $\mathbf{X}$ ,  $\{S_n^*(T_n^*) \leq \beta \hat{S}_n(T_n^*)\} \subset \{S_n^* \leq \beta \hat{S}_n \text{ on } [0, T_n^*)\}$ .  $\square$

LEMMA B.2 (Extension of Lemma 2.7 in Gill, 1983). *For any  $\beta \in (0, 1)$ ,*

- (a)  $P(\hat{H}_n(t-) \geq \beta H_n(t-) \text{ for all } t \leq \hat{T}_n) \geq 1 - \frac{\epsilon}{\beta} \exp(-1/\beta),$
- (b)  $P(H_n^*(t-) \geq \beta \hat{H}_n(t-) \text{ for all } t \leq T_n^* \mid \mathbf{X}) \geq 1 - \frac{\epsilon}{\beta} \exp(-1/\beta) \text{ almost surely.}$

*Proof of (a).* As pointed out by Gill (1983), the assertion follows from the inequality for the uniform distribution in Remark 1(ii) of Wellner (1978). By using quantile transformations, his inequality can be shown to hold for random variables having an arbitrary, even discontinuous distribution function.

*Proof of (b).* Fix  $X_i(\omega), \delta_i(\omega), i = 1, \dots, n$ . Since  $H$  in part (a) is allowed to have discontinuities, (b) follows from (a) for each  $\omega$ .  $\square$

Let  $a, b \in D[0, \tau]$  be two (stochastic) jump processes, i.e. processes being constant between two discontinuities. If  $b$  has bounded variation, we define the integral of  $a$  with respect to  $b$  via

$$\int_0^s a db = \sum a(t) \Delta b(t), \quad s \in (0, \tau],$$

where the sum is over all discontinuities of  $b$  inside the interval  $(0, s]$ . If  $a$  has bounded variation, we define the above integral via integration by parts:  $\int_0^s a db = a(s)b(s) - a(0)b(0) - \int_0^s b_- da$ .

LEMMA B.3 (Adaptation of Lemma 2.9 in Gill, 1983). *Let  $h \in D[0, \tau]$  be a non-negative and non-increasing jump process such that  $h(0) = 1$  and let  $Z \in D[0, \tau]$  be a jump process which is zero at time zero. Then for all  $t \leq \tau$ ,*

$$\sup_{s \in [0, t]} h(s) |Z(s)| \leq 2 \sup_{s \in [0, t]} \left| \int_0^s h(u) dZ(u) \right|.$$

*Proof.* The original proof of Lemma 2.9 in Gill (1983) still applies for the most part with the assumptions of this lemma. For the sake of completeness, we present the whole proof.

Let  $U(t) = \int_0^t h(s) dZ(s)$  with a  $t \leq \tau$  such that  $h(t) > 0$ . Then

$$\begin{aligned} Z(t) &= \int_0^t \frac{dU(s)}{h(s)} = \frac{U(t)}{h(t)} - \int_0^t U(s-) d\left(\frac{1}{h(s)}\right) \\ &= \int_0^t (U(t) - U(s-)) d\left(\frac{1}{h(s)}\right) + \frac{U(t)}{h(0)}. \end{aligned}$$

Thus, following the lines of the original proof,

$$\begin{aligned} |h(t)Z(t)| &\leq \left| \int_0^t (U(t) - U(s-)) d\left(\frac{h(t)}{h(s)}\right) \right| + |U(t)|h(t) \\ &\leq 2 \sup_{0 < s \leq t} |U(s)| \left(1 - \frac{h(t)}{h(0)}\right) + \sup_{0 < s \leq t} |U(s)|h(t) \leq 2 \sup_{0 < s \leq t} |U(s)|. \end{aligned}$$

□

## C Bootstrap Version of the Truncation Technique for Weak Convergence

The following lemma is a conditional variant of Theorem 3.2 in Billingsley (1999). Let  $\rho$  be the modified Skorohod metric  $J_1$  on  $D[0, \tau]$  as in Billingsley (1999), i.e.  $\rho(f, g) = \inf_{\lambda \in \Lambda} (\|\lambda\|^o \vee \sup_{t \in [0, \tau]} |f(t) - g(\lambda(t))|)$ , where  $\Lambda$  is the collection of non-decreasing functions onto  $[0, \tau]$  and  $\|\lambda\|^o = \sup_{s \neq t} \left| \log \frac{\lambda(s) - \lambda(t)}{s - t} \right|$ . For an application in the proof of Theorem 1, note that  $\rho(f, g) \leq \sup_{t \in [0, \tau]} |f(t) - g(t)|$ .

LEMMA C.1. *Let  $X : (\Omega, \mathcal{A}, P) \rightarrow (D[0, \tau], \rho)$  be a stochastic process and let the sequences of stochastic processes  $X_{un}$  and  $X_n$  satisfy the following convergences given a  $\sigma$ -algebra  $\mathcal{C}$ :*

- (a)  $X_{un} \xrightarrow{d} Z_u$  given  $\mathcal{C}$  in probability as  $n \rightarrow \infty$  for every fixed  $u$ ,
- (b)  $Z_u \xrightarrow{d} X$  given  $\mathcal{C}$  in probability as  $u \rightarrow \infty$ ,
- (c) for all  $\varepsilon > 0$  and for each subsequence  $(n') \subset (n)$  there exists another subsequence  $(n'') \subset (n')$  such that

$$\lim_{u \rightarrow \infty} \limsup_{n'' \rightarrow \infty} P(\rho(X_{un''}, X_{n''}) > \varepsilon \mid \mathcal{C}) = 0 \quad \text{almost surely.}$$

Then,  $X_n \xrightarrow{d} X$  given  $\mathcal{C}$  in probability as  $n \rightarrow \infty$ .

*Proof.* Choose a sequence  $\varepsilon_m \downarrow 0$ . Let  $(n') \subset (n)$  be an arbitrary subsequence and choose subsequences  $(n''(\varepsilon_m)) \subset (n')$  and  $(u') \subset (u)$  such that (a) and (b) hold almost surely and also such that (c) holds along these subsequences. Replace  $(n''(\varepsilon_m))$  by their diagonal sequence  $(n'')$  ensuring (c) simultaneously for all  $\varepsilon_m$ . Let  $F \subset D[0, \tau]$  be a closed subset and let  $F_{\varepsilon_m} = \{f \in D[0, \tau] : \rho(f, F) \leq \varepsilon_m\}$  be its closed  $\varepsilon_m$ -enlargement. We proceed as in the proof of Theorem 3.2 in Billingsley (1999) whereas all inequalities now hold almost surely.

$$P(X_{n''} \in F \mid \mathcal{C}) \leq P(X_{u'n''} \in F_{\varepsilon_m} \mid \mathcal{C}) + P(\rho(X_{u'n''}, X_{n''}) > \varepsilon_m \mid \mathcal{C}).$$

The Portmanteau theorem in combination with (a) yields

$$\limsup_{n'' \rightarrow \infty} P(X_{n''} \in F \mid \mathcal{C}) \leq P(Z_{u'} \in F_{\varepsilon_m} \mid \mathcal{C}) + \limsup_{n'' \rightarrow \infty} P(\rho(X_{u'n''}, X_{n''}) > \varepsilon_m \mid \mathcal{C}).$$

Condition (b) and another application of the Portmanteau theorem imply that

$$\limsup_{n'' \rightarrow \infty} P(X_{n''} \in F \mid \mathcal{C}) \leq P(X \in F_{\varepsilon_m} \mid \mathcal{C}).$$

Let  $m \rightarrow \infty$  to deduce  $\limsup_{n'' \rightarrow \infty} P(X_{n''} \in F \mid \mathcal{C}) \leq P(X \in F \mid \mathcal{C})$  almost surely. Thus, a final application of Portmanteau theorem as well as the subsequence principle lead to the conclusion that  $X_n \xrightarrow{d} X$  given  $\mathcal{C}$  in probability. □



## References

- M. G. Akritas. Bootstrapping the Kaplan-Meier Estimator. *Journal of the American Statistical Association*, 81 (396):1032–1038, 1986.
- M. G. Akritas and E. Brunner. Nonparametric Methods for Factorial Designs with Censored Data. *Journal of the American Statistical Association*, 92(438):568–576, 1997.
- A. Allignol, J. Beyersmann, T. Gerds, and A. Latouche. A competing risks approach for nonparametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Analysis*, 20(4):495–513, 2014.
- P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer, New York, 1993.
- P. Billingsley. *Convergence of probability measures*. Wiley, New York, second edition, 1999.
- D. Dobler. *Nonparametric inference procedures for multi-state Markovian models with applications to incomplete life science data*. PhD thesis, Universität Ulm, Deutschland, 2016.
- D. Dobler and M. Pauly. Resampling-based Confidence Intervals and Tests for the Concordance Index and the Win Ratio. *Preprint arXiv:1605.04729*, 2016.
- B. Efron. Censored Data and the Bootstrap. *Journal of the American Statistical Association*, 76(374):312–319, 1981.
- R. D. Gill. Censoring and Stochastic Integrals. *Mathematical Centre Tracts 124*, Amsterdam: Mathematisch Centrum, 1980.
- R. D. Gill. Large Sample Behaviour of the Product-Limit Estimator on the Whole Line. *The Annals of Statistics*, pages 49–58, 1983.
- R. D. Gill. Non- and Semi-Parametric Maximum Likelihood Estimators and the von Mises Method (Part 1)[with Discussion and Reply]. *Scandinavian Journal of Statistics*, 16(2):97–128, 1989.
- M. K. Grand and H. Putter. Regression models for expected length of stay. *Statistics in Medicine*, 2015.
- L. Horvath and B. Yandell. Convergence Rates for the Bootstrapped Product-Limit Process. *The Annals of Statistics*, pages 1155–1173, 1987.
- A. Janssen and T. Pauls. How do bootstrap and permutation tests work? *The Annals of Statistics*, 31(3):768–806, 2003.
- S.-H. Lo and K. Singh. The Product-Limit Estimator and the Bootstrap: Some Asymptotic Representations. *Probability Theory and Related Fields*, 71(3):455–465, 1986.
- I. Meilijson. Limiting Properties of the Mean Residual Lifetime Function. *The Annals of Mathematical Statistics*, 43(1):354–357, 1972.
- S. J. Pocock, C. A. Ariti, T. J. Collier, and D. Wang. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European heart journal*, 33(2):176–182, 2012.
- D. Pollard. *Convergence of Stochastic Processes*. Springer, New York, 1984.
- D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer, third edition, 1999.
- W. Stute and J.-L. Wang. The Strong Law under Random Censorship. *The Annals of Statistics*, 21(3):1591–1607, 1993.

- P. Tattar and H. Vaman. Extension of the Harrington-Fleming tests to multistate models. *Sankhya B*, 74(1):1–14, 2012.
- S.-M. Tse. Lorenz curve for truncated and censored data. *Annals of the Institute of Statistical Mathematics*, 58(4): 675–686, 2006.
- A. W. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, 1996.
- J.-G. Wang. A Note on the Uniform Consistency of the Kaplan-Meier Estimator. *The Annals of Statistics*, pages 1313–1316, 1987.
- J. A. Wellner. Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 45(1):73–88, 1978.
- Z. Ying. A note on the asymptotic properties of the product-limit estimator on the whole line. *Statistics & Probability Letters*, 7(4):311–314, 1989.